

IN THE UNITED STATES DISTRICT COURT
FOR THE SOUTHERN DISTRICT OF TEXAS
CORPUS CHRISTI DIVISION

MARC VEASEY, *et al.*,

Plaintiffs,

v.

Civil Action No. 2:13-cv-193 (NGR)

RICK PERRY, *et al.*,

Defendants.

UNITED STATES OF AMERICA,

Plaintiff,

TEXAS LEAGUE OF YOUNG VOTERS
EDUCATION FUND, *et al.*,

Plaintiff-Intervenors,

TEXAS ASSOCIATION OF HISPANIC
COUNTY JUDGES AND COUNTY
COMMISSIONERS, *et al.*,

Plaintiff-Intervenors,

v.

Civil Action No. 2:13-cv-263 (NGR)

STATE OF TEXAS, *et al.*,

Defendants.

TEXAS STATE CONFERENCE OF NAACP
BRANCHES, *et al.*,

Plaintiffs,

v.

JOHN STEEN, *et al.*,

Defendants.

Civil Action No. 2:13-cv-291 (NGR)

BELINDA ORTIZ, *et al.*,

Plaintiffs,

v.

Civil Action No. 2:13-cv-348 (NGR)

STATE OF TEXAS, *et al.*,

Defendants

DECLARATION OF MICHAEL MIMS

I, Michael Mims, pursuant to 28 U.S.C. § 1746, declare as follows:

1. I currently serve as Program Manager for HealthCare Identity Management for the Veterans Health Administration of the U.S. Department of Veterans Affairs. I have served in this capacity for April 2012. My duties include the maintenance and validation of the identity data stored within the Master Veteran Index.
2. I have personal knowledge of the information contained in this declaration based upon my work for the Veterans Health Administration of the United States Department of Veterans Affairs (“VHA”), and my personal involvement in supervision of the database comparison described below.

3. On March 11, 2014, my agency received an encrypted hard drive containing data regarding Texas registered voters from the United States Department of Justice in the form of a 7 gigabyte .csv file containing 13,564,420 records for Texas registered voters. This file was corrupted in the process of loading the data and so an identical dataset was delivered to VHA on March 26, 2014. VHA was able to load all 13,564,420 records.

4. The Texas data that VHA received contained the following fields:

- Vuid
- last_name
- first_name
- middle_name
- ssn
- ssn4_str
- gender_code
- official_id
- dl_id_str
- date_of_birth
- voter_status_code
- block_number
- street_name
- city
- state
- zip_code
- zip_str
- mail_adrs_1
- last_fix
- multi_lastname1
- multi_lastname2
- first_fix
- middle_initial
- fem_str
- dob_str
- zip5_res_alt
- str_num_res_alt
- AGDN
- AGDN_TAG
- AGDN_unique
- AGDNlast
- AGDNlast_TAG
- AGDNlast_unique
- AGD
- AGD_TAG
- AGD_unique
- ADN
- ADN_TAG
- ADN_unique
- AGN
- AGN_TAG
- AGN_unique
- GDN
- GDN_TAG
- GDN_unique
- DLN
- DLN_TAG
- DLN_unique
- unique_count
- FMLD
- FMLD_TAG
- FMLD_unique
- SDZ
- SDZ_TAG
- SDZ_unique
- SND
- SND_TAG
- SND_unique
- FML1D
- FML1D_TAG
- FML1D_unique
- FML2D

- | | |
|-----------------|------|
| - FML2D_TAG | - K2 |
| - FML2D_unique | - K3 |
| - unique_count2 | - L2 |
| - var_fill | - L3 |

5. From the fields listed above, VHA loaded the following variables that were specifically used in completing the matching process:

- | | |
|------------|------------------|
| - vuid | - last_fix |
| - GDN | - first_fix |
| - FMLD | - middle_initial |
| - FML1D | - dob_str |
| - FML2D | - SDZ |
| - AGDN | - SND |
| - AGDNlast | - var_fill |
| - AGD | - K2 |
| - ADN | - K3 |
| - AGN | - L2 |
| - SSN | - L3 |
| - ssn4_str | |

6. VHA executed the data preparation and comparison steps, attached as Ex. A, that were provided by the United States Department of Justice.

7. Through this process, VHA attempted to match particular combinations of identifying information for Texas registered voters (*e.g.*, first and last name, gender, and date of birth) with the same or related combinations of identifying information with respect to holders of a Veteran Identification Card. For each particular combination, VHA identified instances where the identifying information for a Texas registered voter matched with the same or related combinations of identifying information for one or more holders of a Veteran Identification Card. The frequencies of any missing values in the underlying VHA dataset are provided in Ex. B.

8. For each sweep through the relevant VHA dataset, we appended a column to the Texas data that VHA loaded to indicate, on a record-by-record basis, the output of the database

comparisons on all of the requested combinations where there was one or more matches. In total, out of the 13,564,420 records available to be matched, VHA found matches on one or more combinations for 299,320 voter records.

9. On April 30, 2014, VHA transferred responsive data to the Department of Justice, in the form of a 117 megabyte text (.txt) file, with vertical bar delimiters. This file contained the match results for each of the 299,320 records where VHA found one or more matches. The voter file records where no matches were found were not returned.
10. The results of each sweep, as described in Ex. A, can be found in the following columns of the data VHA returned to the Department of Justice:
 - a. “VHA_USA” contains results of the Stage 1, Step 3.1.1, Combination A match;
 - b. “VHA_USB” contains results of the Stage 1, Step 3.1.2, Combination B match;
 - c. “VHA_USC”contains results of the Stage 1, Step 3.1.2, Combination C match;
 - d. “VHA_USD” contains results of the Stage 1, Step 3.1.2, Combination D match;
 - e. “VHA_USE” contains results of the Stage 1, Step 3.1.2, Combination E match;
 - f. “VHA_USF” contains results of the Stage 1, Step 3.1.2, Combination F match;
 - g. “VHA_USG” contains results of the Stage 2, Step 3.2.1, Combination G match;
 - h. “VHA_USH” contains results of the Stage 2, Step 3.2.2, Combination H match;
 - i. “VHA_USI” contains results of the Stage 2, Step 3.2.2, Combination I match;
 - j. “VHA_USSSN” contains results of the Stage 2, Step 3.2.2, nine-digit social security number match;
 - k. “VHA_USK2”contains results of the Stage 2, Step 3.2.3, Combination K to G match;
 - l. “VHA_USK” contains results of the Stage 2, Step 3.2.3, Combination K match;
 - m. “VHA_USK3” contains results of the Stage 2, Step 3.2.3, Combination K to L match;

- n. “VHA_USL2” contains results of the Stage 2, Step 3.2.4, Combination L to G match;
- o. “VHA_USL” contains results of the Stage 2, Step 3.2.4, Combination L match;
- p. “VHA_USL3” contains results of the Stage 2, Step 3.2.4, Combination L to K match;
- q. “VHA_USF_NW” contains results of the Stage 3, Step 3.3.1, Combination F match;
- r. “VHA_USG_NW” contains results of the Stage 3, Step 3.3.2, Combination G match;
- s. “VHA_USI_NW” contains results of the Stage 3, Step 3.3.2, Combination I match;
- t. “VHA_USSSN_NW” contains results of the Stage 3, Step 3.3.2, nine-digit social security number match;
- u. “VHA_USK2_NW” contains results of the Stage 3, Step 3.3.3, Combination K to G match;
- v. “VHA_USK_NW” contains results of the Stage 2, Step 3.3.3, Combination K match;
- w. “VHA_USK3_NW” contains results of the Stage 2, Step 3.3.3, Combination K to L match;
- x. “VHA_USL2_NW” contains results of the Stage 3, Step 3.3.4, Combination L to G match;
- y. “VHA_USL_NW” contains results of the Stage 3, Step 3.3.4, Combination L match;
- z. “VHA_USL3_NW” contains results of the Stage 3, Step 3.3.4, Combination L to K match;
- aa. “texas1” contains results of the Stage 4, Step 3.4.1, Sweep 1 match;
- bb. “texas2” contains results of the Stage 4, Step 3.4.2, Sweep 2 match;
- cc. “texas3” contains results of the Stage 4, Step 3.4.3, Sweep 3 match; and
- dd. “texas4” contains results of the Stage 4, Step 3.4.4, Sweep 4 match.

11. No other matching comparisons between the Texas data and data from VHA were undertaken beyond those set forth in Ex. A.

I declare under penalty of perjury that the foregoing is true and correct.

Executed on May 27, 2014.

Mildred Mire

Exhibit A

United States v. Texas: Federal Agency Algorithm Summary

This document summarizes the database matches that both the United States and all other plaintiffs and the State of Texas have requested from Federal Agencies as part of the *Veasey v. Perry/United States v. Texas* litigation (S.D. Tex).

The matching process proceeds in three parts, which are explained in detail below. *First*, databases are prepared and standardized. *Second*, identifier values are constructed by combining multiple individual fields. *Third*, the United States' one-to-many matches and the State of Texas's many-to-many matches are conducted between databases.

PART I: DATABASE PREPARATION

Stage 1: Extraction of Available Data from Federal Identification & Disability Databases

- Step 1.1.1:** Extract complete name into separate first name, middle name, and last name fields.
- Step 1.1.2:** Extract date of birth.
- Step 1.1.3:** Extract gender.
- Step 1.1.4:** Extract residential address and mailing address.
- Step 1.1.5:** Extract social security number.
- Step 1.1.6:** Extract Texas driver license number (only if present in Federal database).

Stage 2: Separate Valid Identification and Disability Records

- Step 1.2.1:** Remove records from identification database extracts that indicate that an ID has been revoked or has expired more than 60 days before the date of the TEAM database snapshot (which is January 15, 2014).

Exhibit A

Step 1.2.2: Remove records from disability database extracts that do not indicate current disability status or indicate a Veterans Administration disability rating of less than 50%.

Stage 3: Diagnostics

Step 1.3.1: Report the frequency of missing values for each field.

Step 1.3.2: Report the frequencies of invalid Social Security numbers, such as 111111111 and 123456789.

Step 1.3.3: Report the frequencies of likely invalid dates of birth, such as January 1, 1901 and November 11, 1911.

Stage 4: Standardize Last Name

Step 1.4.1: Remove last name suffixes that are contained within the last name field, rather than a distinct suffix field. *E.g.*, <Smith Jr.> becomes <Smith>.

Step 1.4.2: For last names containing hyphens, populate separate last name fields for all parts of the last name. *E.g.*, the last name <Smith-Jones> would have the value <Smith> entered into a LastName1 field and the value <Jones> entered into a LastName2 field.

Step 1.4.3: Remove spaces, hyphens, periods, and apostrophes from all last name fields and convert all letters to uppercase. *E.g.*, <O'Connor> becomes <OCONNOR> and <Smith-Jones> becomes <SMITHJONES>.

Step 1.4.4: Code all missing values as blank fields.

Exhibit A

Stage 5: Standardize First Name and Middle Name

- Step 1.5.1:** Remove spaces, hyphens, periods, and apostrophes from the first name field and convert all letters to uppercase. *E.g.*, <Jean-Paul> becomes <JEANPAUL>.
- Step 1.5.2:** Parse the first letter of the middle name (if available) and use it to populate a middle initial field. *E.g.*, <John> would yield <J>¹.
- Step 1.5.3:** Code all missing values as blank fields.

Stage 6: Standardize Date of Birth

- Step 1.6.1:** Convert the date of birth to an eight-digit string of MMDDYYYY.
- Step 1.6.2:** Code all missing values as blank fields.

Stage 7: Standardize Gender

- Step 1.7.1:** Code gender as a string of 1 for females and 0 for males.
- Step 1.7.2:** Fill missing gender values using the most common gender value for the first name associated with a record. *E.g.*, if 99% of records with first name <JOHN> are listed as male, assign the male identifier to all records with first name <JOHN> and no listed gender.
- Step 1.7.3:** If missing values remain, code all missing values as blank fields.

¹ The U.S. Department of State does not maintain a separate field for middle names in its database of U.S. Passport and Passport Card holders. Instead, both first and middle name may be stored in the first name field. For this database, the following rule will be applied: treat the first word in the first name field as the first name, and treat the first letter following the first space as the middle initial.

Exhibit A

Stage 8: Standardize Address

- Step 1.8.1:** Convert the residential ZIP code to a string if it is stored as a numeric field.
- Step 1.8.2:** Where the residential address ZIP code is blank, populate that field with the value in the mailing address ZIP code field, if available.²
- Step 1.8.3:** Truncate the residential ZIP code field to the first five digits. *E.g.*, <77777-1234> becomes <77777>.
- Step 1.8.4:** Where the residential address field is blank, populate that field with the value in the mailing address field, if available.
- Step 1.8.5:** Where address field containing street address begins with a street number, isolate the street number. *E.g.*, <123 Main Street> becomes <123>.
- Step 1.8.6:** Where the address field begins with recognized strings indicating a mail box, eliminate strings to isolate the box number. *E.g.*, <PO Box 444> becomes <444>.
- Step 1.8.7:** If missing values remain, code all missing values as blank fields.

Stage 9: Standardize Social Security Number

- Step 1.9.1:** Convert the social security number to a string if it is stored as a numeric field.
- Step 1.9.2:** Using full social security number, check for invalid SSNs. In the case of invalid SSNs, code as missing. *E.g.*, <123456789> becomes <>.

² For purposes of this database matching protocol, the only address fields utilized with respect to data regarding U.S. Passports and U.S. Passport Cards are those regarding mailing addresses.

Exhibit A

Step 1.9.3: Extract the last four digits of full social security number as a four-character string and use them to populate a separate SSN4 field.

Step 1.9.4: Code all missing values as blank fields.

PART II: DATABASE PREPARATION

Stage 1: Construct Primary Identifier Variables for United States' One-to-Many Sweeps

Step 2.1.1: Create Combination A: First Name + Last Name + Gender + DOB + Residential ZIP + Residential Street Number. *E.g.*, the separate fields <JEAN>, <SMITH>, <0>, <01011950>, <77777>, and <123> are combined to a single field <JEANSMITH00101195077777123>.³

Step 2.1.2: Create Combination B: Last Name + Gender + DOB + Residential ZIP + Residential Street Number.

Step 2.1.3: Create Combination C: Gender + Date of Birth + Residential ZIP + Residential Street Number.

Step 2.1.4: Create Combination D: First Name + Last Name + Date of Birth + Residential ZIP + Residential Street Number.

Step 2.1.5: Create Combination E: First Name + Last Name + Gender + Residential ZIP + Residential Street Number.

Step 2.1.6: Create Combination F: First Name + Last Name + Gender + DOB.

Step 2.1.7: Create Combination M: Texas Driver License Number (where available).

³ For the U.S. Department of State only, the name portion of any combination is truncated if it is more than 32 characters long.

Exhibit A

Stage 2: Construct Secondary Identifier Variables for United States' One-to-Many Sweeps

- Step 2.2.1:** Create Combination G: First Name + Middle Initial + Last Name + Date of Birth.⁴
- Step 2.2.2:** Create Combination H: SSN4 + Date of Birth + Residential ZIP.
- Step 2.2.3:** Create Combination I: SSN4 + First Name + Last Name + Date of Birth.
- Step 2.2.4:** Create Combination K: First Name + Last Name 1 + Middle Initial + Date of Birth.
- Step 2.2.5:** Create Combination L: First Name + Last Name 2 + Middle Initial + Date of Birth.
- Step 2.2.6:** Full Social Security Number.

Stage 3: Construct Identifiers Used Only For Texas's Many-to-Many Sweeps

- Step 2.3.1:** Create Combination for Texas's Sweep 1: SSN4 + Last Name + DOB.
- Step 2.3.2:** Create Combination for Texas's Sweep 3: First Name + Last Name + DOB
- Step 2.3.3:** Create Combination for Texas's Sweep 4: First Name + Middle Initial + Last Name + DOB

Note: Combinations for Texas's Sweeps 1 and 3 do not already exist as pre-made fields in the TEAM database extract but instead must be created from the underlying TEAM database fields, in addition to being constructed on the Federal database side. Texas's Sweep 4 is equivalent to

⁴ Only for the State Department, create three further variations of Combination G created using the State Department's "LFMName" field which contains Last, First, and Middle Names, in that order, truncated to a maximum length of 32 characters. Combination G1 is DOB + LFMName; Combination G2 is DOB + First two words of LFMName; and Combination G3 is DOB + First two words of LFMName + First character of third word of LFMName.

Exhibit A

the combination for the United States' Combination G. Texas's Sweep 2 is on full 9 social security number.⁵

Stage 4: Establish Identifier Uniqueness For Combinations A - L

Step 2.4.1: Generate a field that establishes the uniqueness of each identifier variable.

For federal databases, for each combination A-L, generate a field that establishes uniqueness among only Texas records and a field that establishes uniqueness among nationwide records. *E.g.*, if only one record has the string <JEANSMITH0101195001237777> for Combination A, populate the uniqueness field for Combination A for that record as <1>. If four records have the string <JOHNSMITHA0101950> for Combination G, populate the uniqueness field for Combination G for each of those records as <2>, which indicates any number greater than one.

⁵ For purposes of matching to U.S. Department of State Passport and Passport Card holder data, Texas's Sweep 3 and Sweep 4 are as follows: Sweep 3: Last Name + First Name (restricted to 32 characters) + DOB. Sweep 4: Last Name + First Name + Middle Initial (restricted to 32 characters) + DOB.

Exhibit A

PART III: MATCH DATABASES

Stage 1: United States' Primary One-to-Many Matching Sweeps

- Step 3.1.1:** For each case in which Combination A is unique in the TEAM database, match Combination A against Combination A in the identifier or disability database. For federal databases, use only the subset of records with Texas addresses in the identifier or disability database. Where a match is attempted but no match is found, indicate a zero in the Combination A output field. Where there is a match, indicate the uniqueness of Combination A in the identifier or disability database in the Combination A output field (*e.g.*, in cases where there is one matching record in the Federal database, <1> should be inserted into the Combination A output field, while a <2> should be inserted into the Combination A output field if the TEAM record matched 2 or more records in the Federal database).
- Step 3.1.2:** Use the procedure in Step 3.1.1 to match Combination B, Combination C, Combination D, Combination E, and Combination F in the TEAM database against the equivalent combination field in the identifier or disability database.
- Step 3.1.3:** Use the procedure in Step 3.1.1 to match Combination M in the TEAM database against the equivalent combination field in the identifier databases produced by the State of Texas.

Exhibit A

Stage 2: United States' Secondary One-to-Many Matching Sweeps

- Step 3.2.1:** For each case in which no matches were found in the primary one-to-many matching sweeps (A-F, M), and where Combination G is unique in the TEAM database, match Combination G against Combination G in the identifier or disability database. For federal databases, use only the subset of records with Texas addresses in the identifier or disability database. Where a match is attempted but no match is found, indicate a zero in the Combination G output field. Where there is a match, indicate the uniqueness of Combination G in the identifier or disability database in the Combination G output field (*e.g.*, <1> if a unique match and <2> if matched to more than one record).
- Step 3.2.2:** For each case in which no matches were found in the primary one-to-many matching sweeps (A-F, M), use the procedure in Step 3.2.1 to match Combination H, Combination I, and complete social security number⁶ in the TEAM database against the equivalent combination/field in the identifier or disability database.
- Step 3.2.3:** For each case in which no matches were found in the primary one-to-many matching sweeps (A-F, M), use the procedure in Step 3.2.1 to match Combination K against Combination G, Combination K, and Combination L in the identifier or disability database.

⁶ The full social security number is not created as a separate “combination” as it is its own field stored within the TEAM database under the field name “ssn”.

Exhibit A

Step 3.2.4: For each case in which no matches were found in the primary one-to-many matching sweeps (A-F, M), use the procedure in Step 3.2.1 to match Combination L against Combination G, Combination K, and Combination L in the identifier or disability database.

Stage 3: United States' Nationwide Federal Sweeps

Step 3.3.1: For each case in which no matches were found in the primary and secondary matching sweeps of Texas records in a federal identifier or disability database, and where Combination F is unique, match Combination F against Combination F in the nationwide identifier or disability database. Where a match is attempted but no match is found, indicate a zero in the Combination F nationwide output field. Where there is a match, indicate the uniqueness of Combination F in the identifier or disability database in the Combination F nationwide output field (*e.g.*, <1> if a unique match and <2> if matched to more than one record).

Step 3.3.2: For each case in which no matches were found in the primary and secondary matching sweeps of Texas records in a federal identifier or disability database, use the procedure in Step 3.3.1 to match Combination G, Combination I, and full social security number in the TEAM database against the equivalent combination/field in the nationwide identifier or disability database.

Step 3.3.3: For each case in which no matches were found in the primary and secondary matching sweeps of Texas records in a federal identifier or disability database, use the procedure in Step 3.3.1 to match Combination

Exhibit A

K against Combination G, Combination K, and Combination L in the nationwide identifier or disability database.

Step 3.3.4: For each case in which no matches were found in the primary and secondary matching sweeps of Texas records in a federal identifier or disability database, use the procedure in Step 3.3.1 to match Combination L against Combination G, Combination K, and Combination L in the nationwide identifier or disability database.⁷

Stage 4: Texas' Many-to-Many Nationwide Sweeps

Step 3.4.1 Regardless of whether the combination for Sweep 1 is unique in the TEAM database, match against the equivalent combination in a nationwide search of the Federal database.

⁷ Step 3.3.5 for the State Department only: Match the following Combination G variations from applicable State Department records, first to include only the subset of records with Texas addresses, and then to include all applicable U.S. Passport and Passport Card records nationwide (e.g., without Texas addresses), against the following fields from the TEAM database:

- Combination G1 to DOB + Last_fix + First_fix + Middle_name from the TEAM database;
- Combination G2 to DOB + Last_fix + First_fix from the TEAM database;
- Combination G2 to DOB + Last_fix + First word of First_name from the TEAM database;
- Combination G3 to DOB + Last_fix + First_fix + Middle_Initial from the TEAM database;
- Combination G3 to DOB + Last_fix + First word of First_name + Middle_Initial from the TEAM database; and
- Combination G3 to DOB + Last_fix + First word of First_name + First character of Second word of First_fix from the TEAM database.

Attempt matches for all TEAM records, regardless of whether they matched in any prior sweeps. Indicate <1> if a unique match and <2> if matched to more than one record.

Exhibit A

Step 3.4.2 Regardless of whether full 9 social security number is unique in the TEAM database, for Sweep 2, match against the equivalent field in a nationwide search of the Federal database.

Step 3.4.3 Regardless of whether the combination for Sweep 3 is unique in the TEAM database, match against the equivalent combination in a nationwide search of the Federal database.

Step 3.4.4 Regardless of whether the combination for Sweep 4 (Combination G) is unique in the TEAM database, match against the equivalent combination in a nationwide search of the Federal database.

Note: For each of the Texas many-to-many sweeps:

- Indicate <1> if any TEAM combination matches a single combination in the Federal database
- Indicate <2> if any TEAM combination matches more than one record in the Federal database.
- Indicate <0> if no match is achieved.

Examples:

- If there are two TEAM records that have identical versions of the combination for Sweep 1, and there is one record in the Federal database that matches on this combination, both of the underlying TEAM records will have a matching output of <1> for Sweep 1.
- If there are three TEAM records that have identical versions of the combination for Sweep 3, and there are five records in the Federal database that match on that combination, the three TEAM records will each have a matching output of <2> for Sweep 3.

Exhibit B

Step 1.3 Diagnostics on the VHA Data

There were 5,020,207 records extracted from the VHA data for Veterans with a Veteran Identification Card (VIC) on 2/28/2014. The number of Veterans with a VIC and a Texas residential address was 389,383.

Name

	Number of Records	Percent of Records
Last Name Missing	0	0.00%
First Name Missing	31	0.00%
First Name Only One Character	11,881	0.24%
Middle Initial Missing	661,201	13.17%

Social Security Number

	Number of Records	Percent of Records
SSN Missing	2,481	0.05%
Invalid SSN (begins with '00000')	27	0.00%

Gender

Gender was 100% populated.

Date of Birth

The possibly invalid dates of birth below appear to be default values used in the date of birth (e.g., first day of month, 15th of month, first day of year). This is rare.

	Number of Records	Percent of Records
Incomplete date or outside 1901-1996	77	0.00%
Possibly invalid (01-01-01, etc.)	21	0.00%

Zip Code

	Number of Records	Percent of Records
Zip missing (Most of these were cases with a missing residential address)	132,103	2.63%

Exhibit B

Mailing Zip

There were 5,301 veterans with a mailing address.

	Number of Records	Percent of Records
Mailing Zip missing	20	0.38%